

Quantifying uncertainty on Pareto fronts with Gaussian Process conditional simulations

M. Binois^{a,b,*}, D. Ginsbourger^c, O. Roustant^b

^a*Renault S.A.S., 78084 Guyancourt, France*

^b*École Nationale Supérieure des Mines de Saint-Étienne, UMR CNRS 6158, LIMOS, F-42023 Saint-Étienne, France*

^c*University of Bern, Department of Mathematics and Statistics, Alpeneggstrasse 22, CH-3012 Bern, Switzerland*

Abstract

Multi-objective optimization algorithms aim at finding Pareto-optimal solutions. Recovering Pareto fronts or Pareto sets from a limited number of function evaluations are challenging problems. A popular approach in the case of expensive-to-evaluate functions is to appeal to metamodels. Kriging has been shown efficient as a base for sequential multi-objective optimization, notably through infill sampling criteria balancing exploitation and exploration such as the Expected Hypervolume Improvement. Here we consider Kriging metamodels not only for selecting new points, but as a tool for estimating the whole Pareto Front and quantifying how much uncertainty remains on it at any stage of Kriging-based multi-objective optimization algorithms. Our approach relies on the Gaussian Process interpretation of Kriging, and bases upon conditional simulations. Using concepts from random set theory, we propose to adapt the Vorob'ev expectation and deviation to capture the variability of the set of non-dominated points. Numerical experiments illustrate the potential of the proposed workflow, and it is shown on examples how Gaussian process simulations and the estimated Vorob'ev deviation can be used to monitor the ability of Kriging-based multi-objective optimization algorithms to accurately learn the Pareto front.

Keywords: Multi-objective optimization, Attainment function, Vorob'ev Expectation, Expected Hypervolume Improvement, Kriging

1. Introduction

The interest in Multi-Objective Optimization (MOO) has been growing over the last decades, resulting in the development of numerous dedicated methods, especially in evolutionary MOO [1]. These methods are able to cope with challenging problems occurring when few information about the properties of the objective functions is available (black-box optimization). In some situations, as for example in car crash safety design [2], another difficulty comes from a limited budget of evaluations, because of expensive experiments or high fidelity simulations.

In this context, see e.g. [3] for a review, a common approach is to rely on a surrogate model or metamodel to alleviate the computational costs of the optimization process. In particular, Kriging metamodels have proven to be efficient because they not only give a response surface but also a quantification of prediction uncertainty. In mono-objective optimization, this property has been extensively used following the principles of the Efficient Global Optimization (EGO) algorithm [4] where the Expected Improvement criterion is used to balance between exploitation and exploration. Extensions to MOO have been developed, from scalarization approaches [5, 6] to the use of multi-objective expected improvement criteria such as the Expected Hypervolume Improvement [7].

*Corresponding author. Tel.: +33 4 77 42 02 85;

E-mail addresses: binois@emse.fr (M. Binois), ginsbourger@stat.unibe.ch (D. Ginsbourger), roustant@emse.fr (O. Roustant).

While results about the optimality of solutions from aggregation approaches have been reported (see e.g. [8]), things are more difficult to analyze for MOO and even worse in metamodel based MOO, where an additional source of uncertainty due to surrogate modeling must be taken into account. Besides, the study of the convergence in evolutionary MOO is an ongoing subject of research [9].

Inspired by what has been proposed for Kriging-based excursion sets estimation in Chevalier et al. [10, 11], we propose here to use notions from the theory of random sets [12] for quantifying uncertainty on Pareto fronts, through conditional simulations. The latter are used to estimate the probability that any given point in the objective space is dominated, which is known in performance assessment of multi-objective optimizers as the attainment function [13]. From this we obtain a metamodel-based estimation of the Pareto front using the Vorob'ev expectation [12], with a value of the associated uncertainty: the Vorob'ev deviation. At each stage of the sequential optimization process, an insight is provided to the practitioner about convergence and possibilities of further improvements. Furthermore, with two or three objectives the proposed approach makes it possible to visualize the variability around the estimation of the Pareto front in the objective space.

The paper is organized as follows: Section 2 details notions in MOO and in Gaussian Process Regression upon which the proposed approach is based. In particular Section 2.4 is dedicated to conditional simulations. In Section 3, we propose an original definition of uncertainty using the Vorob'ev expectation and deviation. Finally, Section 4 is dedicated to applications of the proposed methodology to three different test cases, where the potential of the approach to quantify uncertainty and monitor convergence within a sequential MOO algorithm is illustrated.

2. Multi-objective optimization and Gaussian Process Regression

2.1. Notions in MOO

Multi-objective optimizers aim at optimizing (say minimizing) several objectives at once: $f_1(\mathbf{x}), \dots, f_m(\mathbf{x})$ with $\mathbf{x} = (x_1, \dots, x_d)^T$ a vector of decision variables in \mathbf{E} (usually $\mathbf{E} \subset \mathbb{R}^d$) and $f : \mathbf{E} \rightarrow \mathbb{R}^m$ the vector valued function whose coordinates are f_1, \dots, f_m . As the objectives are usually in competition, the existence of an optimal solution minimizing all objectives simultaneously cannot generally be taken for granted. This leads to the definition of compromise solutions following the Pareto dominance: a vector is said to be *dominated* if there exists another vector which is not worse in any objective and better for at least one of them. If a vector is not dominated by any other vector, it is said to be optimal in the Pareto sense.

The set of optimal (or non-dominated) points in \mathbf{E} is called *Pareto set* and the corresponding image by f , composed of non-dominated vectors, is called *Pareto front*. Multi-objective optimization algorithms aim at finding non-dominated objective vectors as close as possible to the true underlying Pareto front, creating a discrete approximation sometimes called a *Pareto front approximation* [14].

2.2. Kriging / Gaussian Process Regression

A common solution to perform optimization under a tight evaluation budget is to appeal to a mathematical surrogate of the objective function. Here we focus on a class of probabilistic metamodels relying on Gaussian random fields. Originating from geostatistics with a technique named *Kriging* [15], predicting with such a metamodel is known in the machine learning community as *Gaussian Process Regression* (GPR) [16]. These Kriging/GPR metamodels have the property of interpolating the observations when noiseless data is considered (deterministic case). Furthermore, due to the probabilistic nature of these metamodels, they also provide a quantification of the prediction uncertainty.

Without loss of generality, here we do not assume a priori any stochastic dependency between the responses f_1, \dots, f_m and we treat them separately since the use of dependent models is significantly more cumbersome and has not been shown to perform better in state of the art studies [17, 18]. Following the settings of Gaussian

Process Regression [19, 20], each of the objective functions f_i is supposed to be a sample path of a random field Y_i :

$$Y_i(\cdot) = \mathbf{g}_i^T(\cdot)\boldsymbol{\beta}_i + Z_i(\cdot) \quad (\text{Universal Kriging})$$

where $\mathbf{g}_i(\cdot)^T$ is a vector of known basis functions, $\boldsymbol{\beta}_i$ a vector of unknown coefficient and $Z_i(\cdot)$ is a zero mean Gaussian process (GP) with given covariance function, or kernel, k_i . With n evaluations at the same locations for the objectives $\{Y_i(\mathbf{x}_1) = y_{i,1}, \dots, Y_i(\mathbf{x}_n) = y_{i,n}, 1 \leq i \leq m\}$ denoted \mathcal{A}_n , the predictor (or Kriging mean) and the prediction covariance (also referred to as Kriging covariance) of Universal Kriging (UK) are expressed as:

$$\begin{aligned} m_{i,n}(\mathbf{x}) &= \mathbf{g}_i(\mathbf{x})^T \hat{\boldsymbol{\beta}}_i + k_{i,n}(\mathbf{x})^T \mathbf{K}_{i,n}^{-1} (\mathbf{y}_{i,n} - \mathbb{G}_{i,n} \hat{\boldsymbol{\beta}}_i) \\ c_{i,n}(\mathbf{x}, \mathbf{x}') &= k_i(\mathbf{x}, \mathbf{x}') - k_{i,n}(\mathbf{x})^T \mathbf{K}_{i,n}^{-1} k_{i,n}(\mathbf{x}') + (\mathbf{g}_i(\mathbf{x})^T - k_{i,n}(\mathbf{x})^T \mathbf{K}_{i,n}^{-1} \mathbb{G}_{i,n}) (\mathbb{G}_{i,n}^T \mathbf{K}_{i,n}^{-1} \mathbb{G}_{i,n})^{-1} (\mathbf{g}_i(\mathbf{x}')^T - k_{i,n}(\mathbf{x}')^T \mathbf{K}_{i,n}^{-1} \mathbb{G}_{i,n})^T \end{aligned}$$

where $\mathbf{y}_{i,n} = (y_{i,1}, \dots, y_{i,n})$, $\mathbf{K}_{i,n} = (k_i(\mathbf{x}_s, \mathbf{x}_t))_{1 \leq s, t \leq n}$, $k_{i,n}(\mathbf{x}) = (k_i(\mathbf{x}, \mathbf{x}_1), \dots, k_i(\mathbf{x}, \mathbf{x}_n))^T$, $\mathbb{G}_{i,n} = (g_i(\mathbf{x}_1)^T, \dots, g_i(\mathbf{x}_n)^T)^T$ and $\hat{\boldsymbol{\beta}}_i = (\mathbb{G}_{i,n}^T \mathbf{K}_{i,n}^{-1} \mathbb{G}_{i,n})^{-1} \mathbb{G}_{i,n}^T \mathbf{K}_{i,n}^{-1} \mathbf{y}_{i,n}$.

Note that from a Bayesian point of view, assuming that the Y_i are Gaussian conditionally on $\boldsymbol{\beta}_i$ and putting an improper uniform prior on $\boldsymbol{\beta}_i$, it is known [21] that the Universal Kriging mean and covariance coincide with the conditional expectation and covariance of Y_i knowing \mathcal{A}_n , respectively: $m_{i,n}(\mathbf{x}) = \mathbb{E}(Y_i(\mathbf{x})|\mathcal{A}_n)$ and $c_{i,n}(\mathbf{x}, \mathbf{x}') = \text{cov}(Y_i(\mathbf{x}), Y_i(\mathbf{x}')|\mathcal{A}_n)$.

The covariance functions are chosen according to prior hypothesis about the unknown functions, such as regularity, sparsity, possible symmetries, etc. [22]. While there exists a variety of admissible covariance functions, the most commonly used are the stationary ‘‘Gaussian’’ and ‘‘Matérn’’ kernels [19]. Maximum likelihood estimation or cross validation techniques [23] are typically employed to estimate values for the kernel hyperparameters [16]. An example of a Kriging model with constant unknown trend and Matérn ($\nu = 5/2$) kernel is depicted in Figure 1a.

2.3. Multi-objective expected improvement

Sequential approaches in MOO aim at adding new observations with a balance between exploration and exploitation. Similar to [4], several extensions of the EGO algorithm have been proposed for MOO. The main idea is to derive criteria in the vein of the *Expected Improvement* by defining a generalization of the notion of improvement for multiple objectives. Popular methods include scalarization approaches like ParEGO [5] or MOEAD-EGO [6] or truly multi-objective methods based on the definition of improvement functions over the current Pareto front \mathcal{P}_n defined by the current observations. Considered improvement functions are respectively based on Euclidean distance [24], Hypervolume [25, 26] or Maximin distance [27, 18] i.e. respectively the distance to the closest point of \mathcal{P}_n , the volume added over \mathcal{P}_n and an axis-wise distance to \mathcal{P}_n .

In the applications of Section 4, we use the Expected Hypervolume Improvement to sequentially add points. This criterion has been successfully applied to problems with limited budget [7], enjoys some beneficial properties [7] and furthermore is related to the concept of attainment function which is of particular importance in what follows.

2.4. Conditional simulations

From the Universal Kriging metamodels presented in Section 2.2, it is possible to generate samples interpolating the available evaluation results, called conditional simulations. They can be generated using a variety of methods, from matrix decomposition to spectral methods, as presented in [28, 29, 30, 31]. Examples of such conditional simulations are displayed in Figure 1b.

They have been applied in mono-objective optimization in [32] as a tool to estimate an information gain when no analytical formula is available, as opposed to the Expected Improvement. Until now, the computation

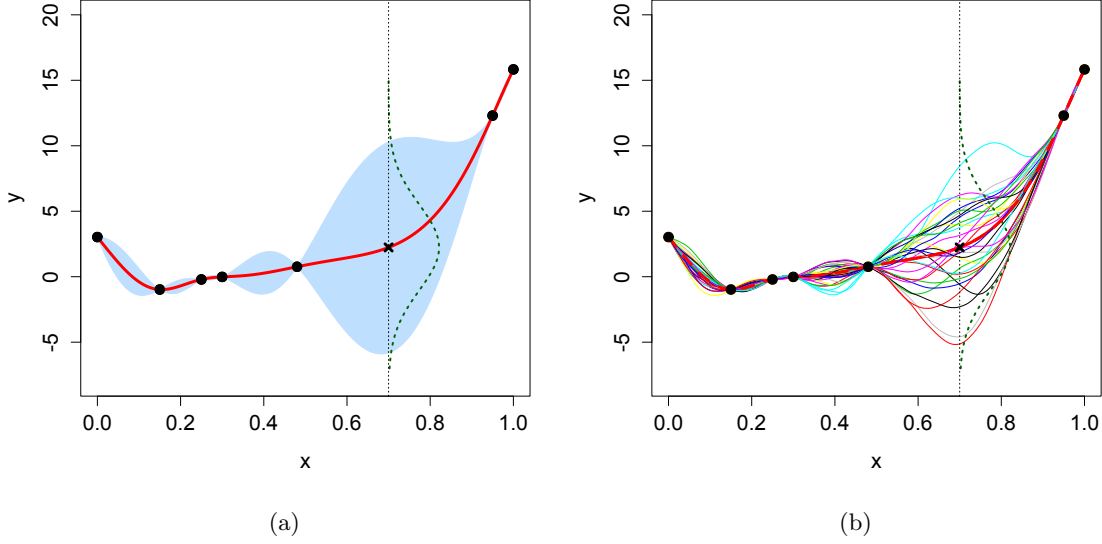


Figure 1: (a) Example of Kriging model based on observations at $n = 7$ locations (black points), with Kriging mean (bold line) and Kriging pointwise 95% prediction intervals (shaded area). The Gaussian predictive distribution for $x = 0.7$ is represented by the vertical dashed line. (b) Conditional simulations (colored thin lines) from the fitted UK metamodel.

of multi-objective Expected Improvement relies either on analytical formulas or on Monte Carlo estimation with draws of the posterior distribution at a single location \mathbf{x} . In contrast, conditional simulations consist in drawing posterior realizations of the unknown function at multiple points, say $\mathbf{E}_p : \{\mathbf{e}_1, \dots, \mathbf{e}_p\} \subset \mathbf{E}$. Since exact methods essentially depend on $p \times p$ covariance matrices, the number of simulation points is typically limited by storage and computational cost. Despite this limitation, conditional simulations (based on matrix decomposition) prove useful for Pareto front estimation, as presented in the next sections.

3. Quantification of uncertainty

In this section, we assume that a Gaussian process model (see Section 2.2) has been estimated for each objective function from a set of n observations \mathcal{A}_n . These models allow us to generate conditional Pareto front realizations and further estimate the uncertainty on the Pareto front with concepts from random sets theory.

3.1. Conditional simulations for MOO: generation of conditional Pareto fronts and corresponding attained sets

Here we use conditional simulations to generate so-called *conditional Pareto fronts* (CPF). The first step is to simulate a finite number (say N) of vector-valued GP conditional simulations $\{\mathbf{Y}_1^{(1)}, \dots, \mathbf{Y}_m^{(1)}\}, \dots, \{\mathbf{Y}_1^{(N)}, \dots, \mathbf{Y}_m^{(N)}\}$ at some simulation points in the design space. Selecting the corresponding non-dominated simulation points and simulated responses then provides conditional Pareto sets and fronts as summarized in Algorithm 1 and illustrated in Figure 2.

Note that what we denote by CPF are actually approximations of conditional Pareto fronts, just like conditional simulations of Gaussian random fields are often approximated realizations relying on a finite number of points. Simulation points can be fixed for all the simulations or changed from one simulation to the other. Fixed simulation points accelerate the simulation generation but they may introduce a bias and lead to missing worthwhile areas. On the other hand, modifying the simulation points increases the computational burden but is more exploratory, which might be an asset in high dimensions. Besides, the procedure used to choose the location of simulation points may impact the results. Accordingly, two sampling strategies are investigated in Section 4.2.

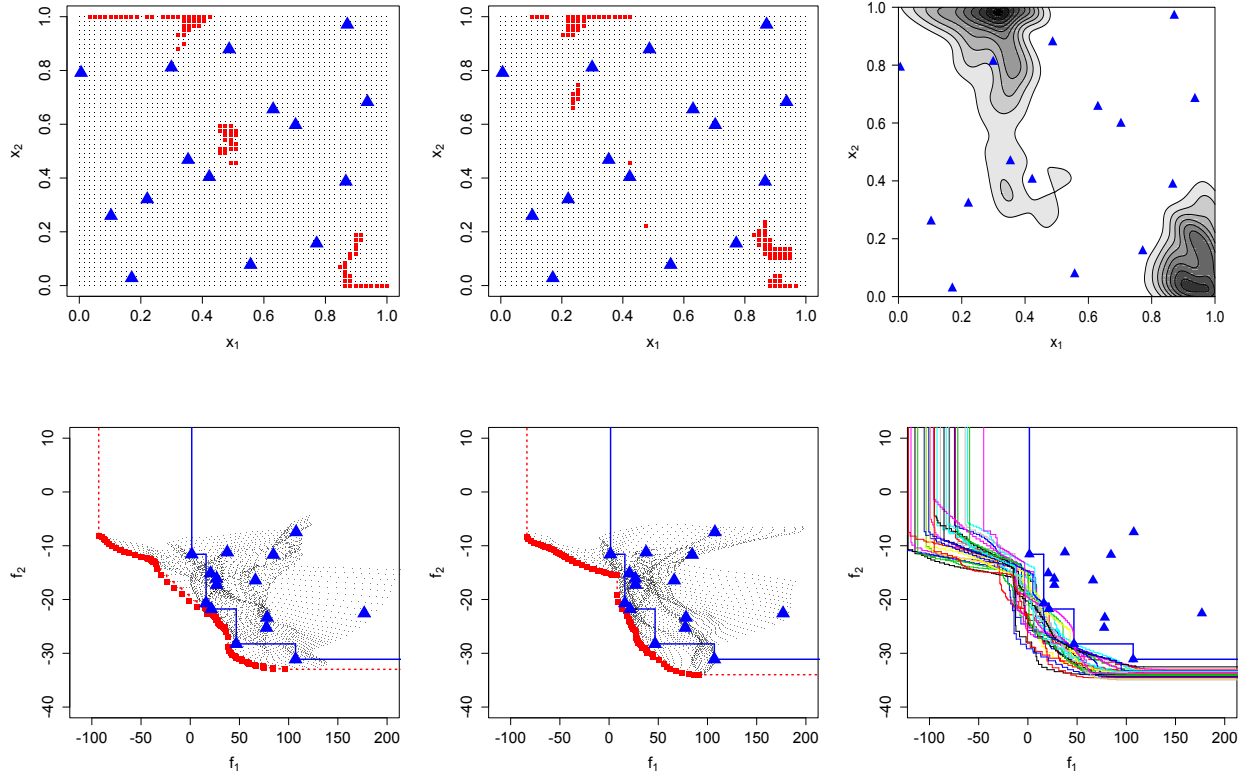


Figure 2: Conditional Pareto sets and fronts corresponding to the GP models Y_1, Y_2 , based on the observations \mathcal{A}_n represented by blue triangles. Left and center: examples of two conditional simulations of Pareto sets (top) and fronts (bottom), where the simulations are performed on a regular 100×100 grid. The simulation points and simulated responses are plotted with dots. The corresponding non-dominated points are represented by red squares. Right: contour plot of the probability density of optimal points in the decision space estimated from 30 conditional simulations (top) and superposition of simulated conditional Pareto fronts (bottom).

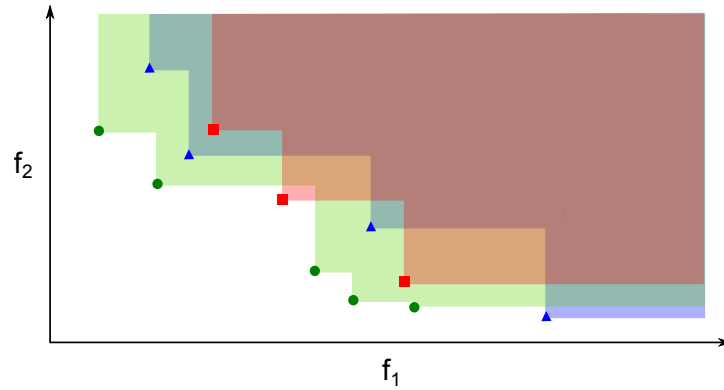


Figure 3: Example of 3 realizations of RNP sets (points, triangles and squares) and the corresponding attained sets (shaded areas).

Algorithm 1 Simulation of N conditional Pareto sets and fronts

```
for  $i = 1, 2, \dots, N$  do
    Choose  $p$  simulation points  $\mathbf{E}_p = \mathbf{e}_1, \dots, \mathbf{e}_p$  in  $\mathbf{E}$  (fixed or different at each iteration).
    for  $j = 1, 2, \dots, m$  do
        Generate a conditional simulation at  $\mathbf{e}_1, \dots, \mathbf{e}_p$  for the  $j^{\text{th}}$  objective:  $\mathbf{Y}_j^{(i)} = (Y_j^{(i)}(\mathbf{e}_1), \dots, Y_j^{(i)}(\mathbf{e}_p))$ .
    end for
    Determine the Pareto set and front of  $\{\mathbf{Y}_1^{(i)}, \dots, \mathbf{Y}_m^{(i)}\}$ .
end for
```

From now on we focus on the use of CPFs since the decision maker is mostly interested in visualizing results in the objective space. Each CPF is composed of non-dominated points in the objective space. They have been considered to assess the performances of MO optimizers [33, 14] under the term Random Non-dominated Point (RNP) sets: sets of random vectors in \mathbb{R}^m , non-dominated with respect to each other and with random finite cardinality (see e.g. [13]). An alternative view is to consider the set of all objective vectors dominated by a realization of an RNP set, called an *attained set*. Realizations of RNP sets and the corresponding attained sets are presented in Figure 3.

In the mono-objective case, GP models provide analytical expressions of the expectation (Kriging mean) and uncertainty (variance of the pointwise prediction). It would be interesting to get their counterpart for the attained sets of simulated CPFs. Nevertheless, defining an expectation and/or an index of variability for sets is not straightforward and requires concepts from random sets theory [12].

3.2. Basics from random sets theory: quantifying uncertainty with the Vorob'ev deviation

Before introducing related notions for CPFs, let us recall some general definitions. Set-valued random elements, in particular *random closed sets* [12] received attention in the probability literature over the last decades. There exist several candidate notions to define the mean of a random closed set, see [12] (Chapter 2). We choose a rather intuitive one, based on the notion of *coverage function*:

Definition 1 (Coverage function). Let \mathcal{Y} be a random closed set on a topological space D (here $D \subset \mathbb{R}^m$ equipped with the topology induced by the Euclidean distance). The coverage function $p_{\mathcal{Y}}$ is defined by $p_{\mathcal{Y}} : x \in D \mapsto \mathbb{P}(x \in \mathcal{Y})$.

This definition has been applied in the Kriging framework to estimate sets of critical input values [10, 11]. It uses the Vorob'ev expectation, based on the upper level sets $\mathcal{Q}_{\beta} = \{z \in D, p_{\mathcal{Y}}(z) \geq \beta\}$, called β -quantiles.

Definition 2 (Vorob'ev expectation and deviation). Denoting by μ the Lebesgue measure on \mathbb{R}^m and assuming that $\mathbb{E}(\mu(\mathcal{Y})) < +\infty$, the Vorob'ev expectation is the β^* -quantile \mathcal{Q}_{β^*} such that $\mathbb{E}(\mu(\mathcal{Y})) = \mu(\mathcal{Q}_{\beta^*})$ if this equation has a solution, and otherwise it is defined from the condition $\mu(\mathcal{Q}_{\beta}) \leq \mathbb{E}(\mu(\mathcal{Y})) \leq \mu(\mathcal{Q}_{\beta^*})$, $\forall \beta > \beta^*$. The associated Vorob'ev deviation is the quantity $\mathbb{E}(\mu(\mathcal{Q}_{\beta^*} \Delta \mathcal{Y}))$, where Δ denotes the symmetric difference between sets, i.e. $\mathcal{Q}_{\beta^*} \Delta \mathcal{Y} = (\mathcal{Q}_{\beta^*} \cup \mathcal{Y}) \setminus (\mathcal{Q}_{\beta^*} \cap \mathcal{Y})$.

The Vorob'ev expectation is a global minimizer of the deviation among all deterministic closed sets with volume equal to the average volume of \mathcal{Y} (see [12] for a proof): for any set M with $\mu(M) = \mathbb{E}(\mu(\mathcal{Y}))$, $\mathbb{E}(\mu(\mathcal{Q}_{\beta^*} \Delta \mathcal{Y})) \leq \mathbb{E}(\mu(M \Delta \mathcal{Y}))$.

3.3. Quantification of uncertainty on Pareto fronts using random set theory

In the MOO literature, the study of distribution location and spread of an attained set \mathcal{X} rely on the attainment function $\alpha_{\mathcal{X}}$ [13]: the probability for a given point in the objective space to be dominated by an RNP set, $\alpha_{\mathcal{X}} = \mathbb{P}(x \in \mathcal{Y})$.

Attained sets are also closed¹ and unbounded subsets in \mathbb{R}^m . Hence, the attained sets obtained with the simulated CPFs can be considered as realizations of a random closed set and are denoted by \mathcal{Y}_i , ($i = 1, \dots, N$). Looking again at Definition 1, one can see that the attainment function is in fact a coverage function. For proofs about the equivalence of the distribution of an RNP set and the corresponding attained set as well as for a definition of the attainment function in terms of coverage function, we refer to [13]. This reference establishes the link between optimization results and random closed sets, in a case where the attainment function is computed from several runs of optimizers. Their comparison is then performed based on statistical hypothesis testing procedures.

In practice the attainment function is estimated by taking the proportion of RNP sets that dominates a given vector in the objective space:

Definition 3 (Empirical attainment function). Given a sample of attained sets $\mathcal{Y}_1, \dots, \mathcal{Y}_N$, the empirical attainment function is defined as: $\hat{\alpha}_N : \mathbb{R}^m \mapsto [0, 1]$, $\hat{\alpha}_N(z) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{z \in \mathcal{Y}_i\}}$ where $\mathbf{1}_{\{z \in \mathcal{Y}_i\}} = 1$ if $z \in \mathcal{Y}_i$, 0 otherwise.

An example of an empirical attainment function is presented in Figure 4, showing where in the objective space there is a high probability to improve on the current Pareto front.

Definition 2 requires that \mathcal{Y} is bounded for its Vorob'ev expectation to exist. Hence it is necessary to define a reference point \mathbf{R} to bound the integration domain. Since the Lebesgue measure of an attained set with respect to the reference point is the hypervolume indicator of the corresponding RNP set, denoted by $I_H(\cdot, \mathbf{R})$, the choice of \mathbf{R} has a similar influence (see e.g. [34]). Unless there is previous knowledge about the range of the objectives, we choose \mathbf{R} as the maximum of each objective reached by the conditional simulations.

The Pareto frontier of the Vorob'ev expectation provides us with an estimate of the Pareto front, as illustrated in Figure 4. The value of the Vorob'ev deviation gives an idea about the variability of the simulated CPF and can be monitored as observations are added, as will be shown in Section 4. The procedure to determine the value β^* corresponding to the Vorob'ev expectation (Vorob'ev threshold) as well as of the Vorob'ev deviation is described in Algorithm 2.

Remark 1. The last equality in Algorithm 2 comes from the following: $\int_{\Omega} \mathbf{1}_{z \in A \Delta B} \mu(dz) = I_{H2}(A, B, \mathbf{R}) + I_{H2}(B, A, \mathbf{R})$ where $I_{H2}(A, B, \mathbf{R})$ is the binary hypervolume indicator, defined for instance in [35]: the volume dominated by A and not by B, i.e. $I_{H2}(A, B, \mathbf{R}) = I_H(A \cup B, \mathbf{R}) - I_H(B, \mathbf{R})$.

The determination of the Vorob'ev threshold β^* requires the volumes of β -quantiles:

$$\mu(\mathcal{Q}_{\beta}) = \int_{\Omega} \mathbf{1}_{\hat{\alpha}_N(z) \geq \beta} \mu(dz)$$

They can be estimated by numerical quadrature, i.e. by computing the values of $\hat{\alpha}_N$ on a grid when there are few objectives or relying on Monte Carlo schemes. The other integrations are performed using hypervolume computation procedures and their complexity is related to the difficulty of computing the hypervolume in general. Hence measuring the uncertainty with the Vorob'ev deviation would be possible with any number of objectives but would require approximating integrals to keep the computations affordable. Note that since we take the objectives separately, simulating with more objectives simply requires computing the additional conditional simulations corresponding to those objectives.

¹As a finite union of closed sets (hyper quadrants).

Algorithm 2 General procedure for estimating the Vorob'ev expectation and deviation

- 1: Generate N CPFs (see Algorithm 1).
- 2: **if** \mathbf{R} is unknown **then** find the extremal values for the objectives over the RNP sets realizations:

$$\mathbf{R} = \left[\max_{i \in (1, \dots, N)} \mathbf{Y}_1^{(i)}, \dots, \max_{i \in (1, \dots, N)} \mathbf{Y}_m^{(i)} \right]$$

- 3: Define the integration domain: $\Omega = \left[\min_{i \in (1, \dots, N)} \mathbf{Y}_1^{(i)}, R_1 \right] \times \dots \times \left[\min_{i \in (1, \dots, N)} \mathbf{Y}_m^{(i)}, R_m \right]$
- 4: Determine the average volume of the attained sets \mathcal{Y}_i :

$$\mathbb{E}(\mu(\mathcal{Y})) \approx \frac{1}{N} \sum_{i=1}^N \int_{\Omega} \mathbf{1}_{\{z \in \mathcal{Y}_i\}} \mu(dz) = \frac{1}{N} \sum_{i=1}^N I_H(\mathcal{Y}_i, \mathbf{R})$$

- 5: Find the value of the Vorob'ev threshold β^* by dichotomy: set $a = 0, b = 1$:

while $b - a < \epsilon$ **do**
 if $\mu\left(\mathcal{Q}_{\frac{a+b}{2}}\right) < \mathbb{E}(\mu(\mathcal{Y}))$ **then** $b = \frac{a+b}{2}$
 else $a = \frac{a+b}{2}$
 end if
end while, $\beta^* = \frac{a+b}{2}$

- 6: Estimate the Vorob'ev deviation:

$$\mathbb{E}(\mu(\mathcal{Q}_{\beta^*} \Delta \mathcal{Y})) \approx \frac{1}{N} \sum_{i=1}^N \int_{\Omega} \mathbf{1}_{(z \in \mathcal{Q}_{\beta^*} \Delta \mathcal{Y}_i)} \mu(dz) = \frac{1}{N} \sum_{i=1}^N (2I_H(\mathcal{Q}_{\beta^*} \cup \mathcal{Y}_i, \mathbf{R}) - I_H(\mathcal{Q}_{\beta^*}, \mathbf{R}) - I_H(\mathcal{Y}_i, \mathbf{R}))$$

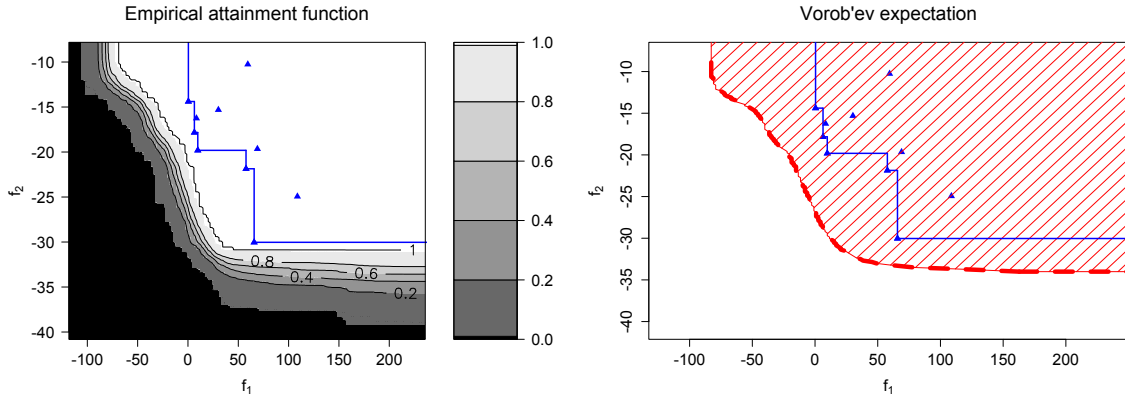


Figure 4: Example of empirical attainment function (level sets, left) and corresponding Vorob'ev expectation (shaded area, right), based on 200 conditional simulations. The estimate of the underlying Pareto front (dashed line) is the Pareto frontier of the Vorob'ev expectation. The ten observations are marked by blue triangles.

Remark 2. [13] proposes the use of the Vorob'ev median ($Q_{0.5}$) if no compact set is chosen for integration. While removing the problem of fixing the reference point, no equivalent of the Vorob'ev deviation seems available in this case.

From a practical point of view, it is also useful for visualization purpose with few objectives to display the superposition of all the symmetric differences by defining an analogue of the attainment function:

Definition 4 (Symmetric-deviation function). The function $\delta_{\mathcal{Y}} : z \in \mathbb{R}^m \mapsto P(z \in \mathcal{Q}_{\beta^*} \Delta \mathcal{Y})$ is called the symmetric-deviation function of \mathcal{Y} .

$\delta_{\mathcal{Y}}$ is the coverage function of $\mathcal{Q}_{\beta^*} \Delta \mathcal{Y}$. It is estimated with the empirical symmetric-deviation function:

$$\hat{\delta}_N(z) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{z \in \mathcal{Q}_{\beta^*} \Delta \mathcal{Y}_i\}}.$$

Figure 5 presents an example of a symmetric difference between two sets and an empirical symmetric-deviation function. This shows the variability around the estimated Pareto front: dark areas indicate regions where the estimation of the Pareto front is not known precisely.

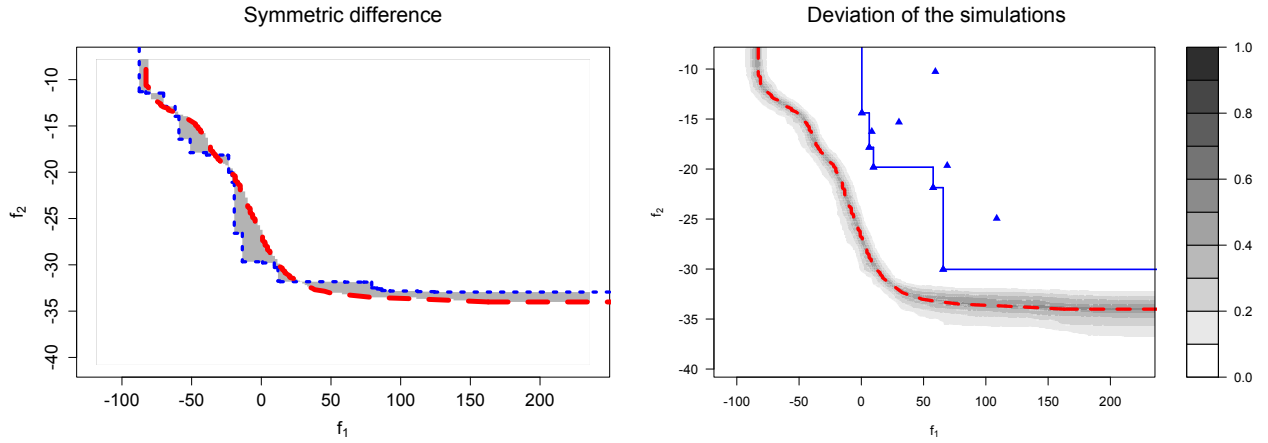


Figure 5: Left: symmetric difference between the Vorob'ev expectation (the level line of the Vorob'ev threshold is represented by the red dashed line) and a simulated CPF's attained set (blue dotted line). Right: illustration of the deviation around the estimated Pareto front corresponding to Figure 4 with an example of empirical symmetric-deviation function (level plot).

4. Application

4.1. Two-dimensional bi-objective test problems

In this section, we illustrate the benefits of the proposed methodology for estimating Pareto fronts. We consider the following two variable, bi-objective optimization problems from the literature:

(P1) The problem presented in [36], which has a convex Pareto front:

$$f_1(\mathbf{x}) = \left(b_2 - \frac{5.1}{4\pi^2} b_1^2 + \frac{5}{\pi} b_1 - 6 \right)^2 + 10 \left[\left(1 - \frac{1}{8\pi} \right) \cos(b_1) + 1 \right]$$

$$f_2(\mathbf{x}) = -\sqrt{(10.5 - b_1)(b_1 + 5.5)(b_2 + 0.5)} - \frac{1}{30} \left(b_2 - \frac{5.1}{4\pi^2} b_1^2 - 6 \right)^2 - \frac{1}{3} \left[\left(1 - \frac{1}{8\pi} \right) \cos(b_1) + 1 \right]$$

where $b_1 = 15x_1 - 5$, $b_2 = 15x_2$ and $x_1, x_2 \in [0, 1]$.

(P2) The ZDT3 problem [37] which has a disconnected Pareto front.

For each example, we start with a set of few observations that allow fitting initial Gaussian process models for the two objective functions. Then we add new points sequentially by maximizing the Expected Hypervolume Improvement, based on the formula detailed in [7]. At each step, the Gaussian process models are updated and their hyperparameters re-estimated. These models are then used to simulate CPFs, from which we compute the estimates of the Vorob'ev mean and the measures of uncertainty: Vorob'ev deviation and symmetric-deviation function. Since the integration domain varies as points are added, the values are displayed divided by the volume of this integration domain. The following test problems are fast to compute, so it is possible to compare the outcome of the proposed workflow to a reference Pareto front by using the volume of the symmetric difference.

The results are presented in Figure 6 and Figure 7, showing the evolution of the estimated Pareto fronts with the corresponding uncertainty around it. For the problem (P1) the sequence is detailed, demonstrating the strength of the proposed approach for giving insights about the uncertainty on the Pareto front. In particular, the uncertainty measures are helpful for choosing a minimal number of observations for approximating the Pareto front: while 10 initial observations may not be enough (Figure 6a) regarding the large symmetric-deviation, adding 10 more observations dramatically reduces the uncertainty (Figure 6c).

The conclusions are similar for problem (P2) Figure 7, where the Pareto front is disconnected, starting this time with 20 observations and sequentially adding ten more observations by Expected Hypervolume Improvement maximization. This example makes clear that the Vorob'ev expectation refers to the area dominated by the Pareto front. When the latter is disconnected, the dominated area's frontier is horizontal in the corresponding parts. The position of the cuts in the Pareto front depends on the model and simulations, resulting in an higher variation around cuts: a small change in the extent of a peak impacts the beginning of the next one (where the symmetric difference volume depends on the size of the disconnection).

One can note that the approximations obtained are dependent on the model accuracy. In Figure 6a the approximation is clearly too optimistic, due to an underestimation of the range parameters of the used Matérn covariance kernel ($\nu = 5/2$). Similarly, for one objective, the expected value of the minimum would be misleading at the beginning.

Finally, we propose the use of the Vorob'ev deviation as a stopping criterion when the Pareto front location is known. An empirical rule could be a threshold on the Vorob'ev deviation (e.g. expected volume of the symmetric difference less than 1% of the integration volume) and detection of stagnation (e.g. under the threshold for several evaluations). On the examples Figures 6d and 7b, by considering the two last evaluations, the result would have been to stop for problem (P1) and continue for problem (P2). Note that this simple criterion would fail if the estimation of hyperparameters is misleading. A more robust version would be to integrate the uncertainty on the hyperparameters estimation in a full Bayesian framework [30]. More sophisticated values could also be derived inspired from [9].

4.2. Additional experiments on conditional simulations

The aforementioned methodology depends on the number and location of simulation points used to obtain the CPFs from the Gaussian process models (cf. Algorithm 1 and first step of Algorithm 2). As a first study, we have compared two sampling strategies: uniform sampling and space-filling sampling relying on a Sobol sequence, again with 2 variables. The objective functions are taken as sample paths of centered Gaussian processes with Matérn covariance kernel ($\nu = 5/2$), with range parameters equal to $0.3/\sqrt{3}$ for f_1 , and $0.5/\sqrt{3}$ for f_2 .

We compute the approximation error over the set of non-dominated points obtained from the two sampling strategies. To compare the results with a reference set obtained with an NSGA-II [38] with archiving, three er-

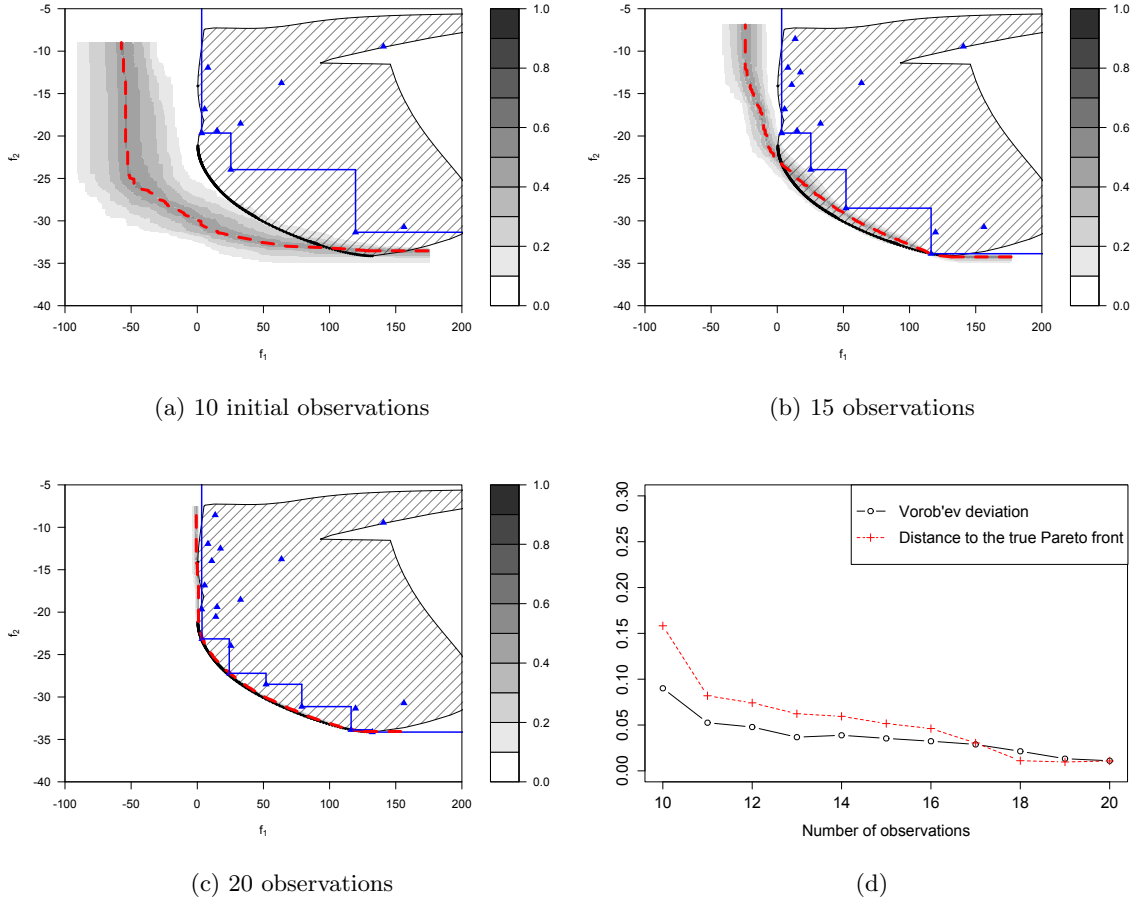


Figure 6: Evolution of the deviation with new observations added using Expected Hypervolume Improvement for Problem (P1). The shaded area represents the image of \mathbf{E} by f with a thicker border for the Pareto front. Observations are marked with blue triangles and the blue solid line represents the current Pareto front. The dashed line is the estimated Pareto front, with the corresponding values of the symmetric-deviation in level plot. Bottom right: evolution of the Vorob'ev deviation scaled by the current integration volume (black solid line with circles) and evolution of the distance to the real Pareto front measured with the volume of the symmetric difference with the estimation (red dotted line with crosses).

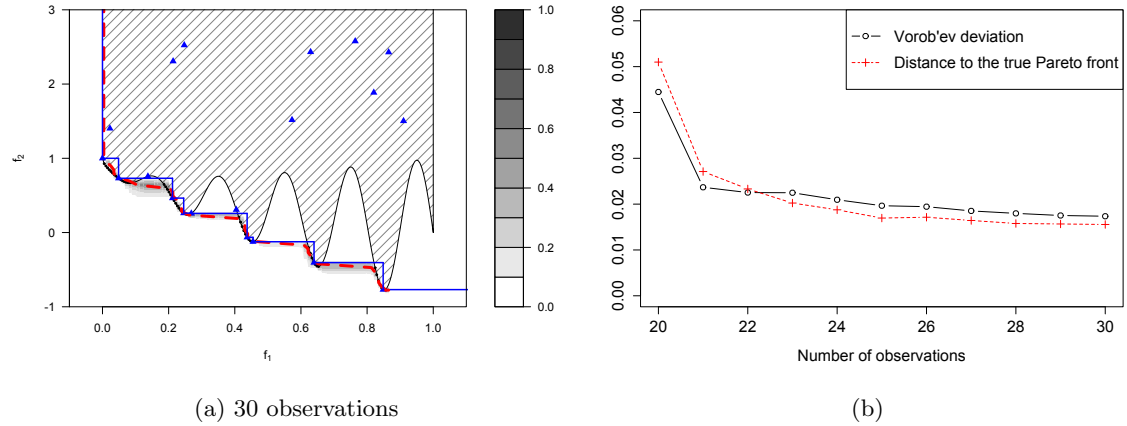


Figure 7: Evolution of the deviation with new observations added using Expected Hypervolume Improvement for Problem (P2). The figure description is the same as in Figure 6.

ror indicators are used: hypervolume difference, epsilon and R2 quality indicators [35]. The tests are repeated one hundred times. The results presented in Figure 8 show that space-filling sampling slightly outperforms uniform sampling to get an accurate estimation of the Pareto front. Additional tests performed showed a similar behavior with 3 variables but no difference for 10 variables, where the number of points considered was too small.

Here, only the impact of the error introduced by discretizing conditional simulations is studied. Given a few hundred to a few thousands of simulation points, it should remain negligible considering a relatively low number of variables, as is usually the case in application examples up to six variables and six objectives [39]. Alternatively, resorting to approximate spectral simulation methods such as the truncated Karhunen-Loève expansion could be considered.

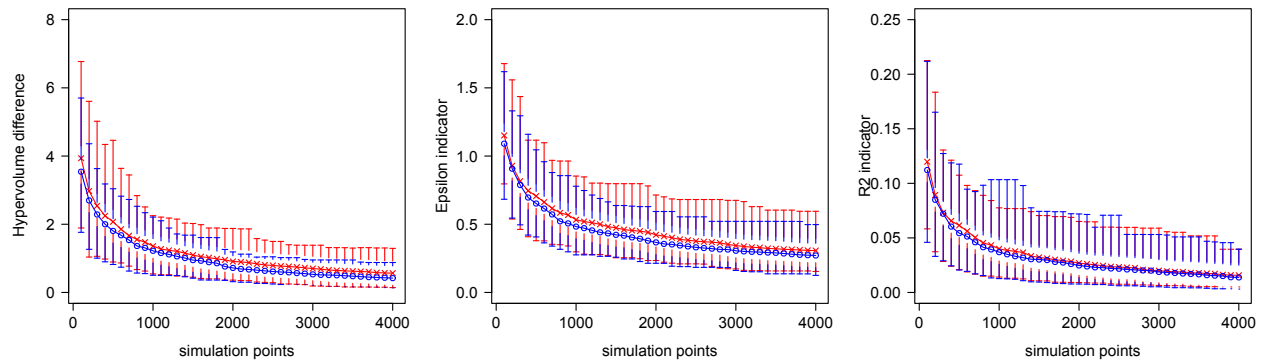


Figure 8: Hypervolume difference, epsilon and R2 quality indicators between random reference Pareto fronts and approximations of them. A hundred conditional simulations for each of the two objectives considered are generated on a bi-dimensional grid with 225 points. Each of them is re-interpolated by Kriging and a reference Pareto front is obtained by applying an NSGA-II [38] with archiving. This reference set is used to compare the quality of approximations of the Pareto front obtained by uniform sampling (red crosses) or with a Sobol sequence (blue points) with respect to the number of simulation points. The error bars indicate the quantiles of level 5% and 95% for the hundred repetitions.

5. Conclusion and perspectives

We presented an original methodology to estimate and visualize the uncertainty of Pareto front approximations, based on Gaussian process conditional simulations. More precisely, the attainment function provides an estimation of the probability of dominating a given point in the objective domain. Then, a global uncertainty measure was defined relying on the theory of random sets through the concept of Vorob’ev deviation. It indicates the confidence of the model on the approximation of the attained set. The last tool is a visualization of the region of confidence for two or three objectives. Application on an higher number of objectives would be feasible, requiring the use of Monte Carlo methods for the computation of the various integrals. As illustrated on two bi-objective problems with convex or disconnected Pareto fronts, these measures can also be used as a basis to define stopping criteria in a sequential framework.

Further work is needed to analyze the different kinds of uncertainty and biases that may occur when applying the proposed methodology. In addition, techniques for simulating efficiently over more points and updating simulations with new observations [40] should be considered, as well as optimization of simulation points locations with re-interpolation or re-simulation [41]. Another direction for future research includes the integration of the proposed uncertainty estimate in a Stepwise Uncertainty Reduction (SUR) strategy [42] as an infill criterion.

Acknowledgments

This work has been conducted within the frame of the ReDice Consortium, gathering industrial (CEA, EDF, IFPEN, IRSN, Renault) and academic (École des Mines de Saint-Étienne, INRIA, and the University of Bern) partners around advanced methods for Computer Experiments. The authors also thanks two anonymous reviewers for their helpful suggestions.

References

- [1] K. Deb, Introduction to evolutionary multiobjective optimization, in: J. Branke, et al. (Eds.), *Multiobjective Optimization*, Vol. 5252 of LNCS, Springer, 2008, pp. 59–96.
- [2] X. Liao, Q. Li, X. Yang, W. Zhang, W. Li, Multiobjective optimization for crash safety design of vehicles using stepwise regression model, *Structural and Multidisciplinary Optimization* 35 (6) (2008) 561–569.
- [3] L. Santana-Quintero, A. Montano, C. Coello, A review of techniques for handling expensive functions in evolutionary multi-objective optimization, *Computational Intelligence in Expensive Optimization Problems* (2010) 29–59.
- [4] D. Jones, M. Schonlau, W. Welch, Efficient global optimization of expensive black-box functions, *Journal of Global Optimization* 13 (4) (1998) 455–492.
- [5] J. Knowles, ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems, *IEEE Transactions on Evolutionary Computation* 10 (1) (2006) 50–66.
- [6] Q. Zhang, W. Liu, E. Tsang, B. Virginas, Expensive multiobjective optimization by MOEA/D with Gaussian process model, *Evolutionary Computation, IEEE Transactions on* 14 (3) (2010) 456–474.
- [7] M. T. Emmerich, A. H. Deutz, J. W. Klinkenberg, Hypervolume-based expected improvement: Monotonicity properties and exact computation, in: *Evolutionary Computation (CEC), 2011 IEEE Congress on*, IEEE, 2011, pp. 2147–2154.
- [8] K. Miettinen, *Nonlinear multiobjective optimization*, Vol. 12, Springer, 1999.
- [9] T. Wagner, H. Trautmann, L. Martí, A taxonomy of online stopping criteria for multi-objective evolutionary algorithms, in: *Evolutionary Multi-Criterion Optimization*, Springer, 2011, pp. 16–30.
- [10] C. Chevalier, D. Ginsbourger, J. Bect, I. Molchanov, Estimating and quantifying uncertainties on level sets using the Vorobev expectation and deviation with Gaussian process models, in: D. Uciniski, A. C. Atkinson, M. Patan (Eds.), *mODa 10 Advances in Model-Oriented Design and Analysis, Contributions to Statistics*, Springer, 2013, pp. 35–43.
- [11] C. Chevalier, Fast uncertainty reduction strategies relying on Gaussian process models, Ph.D. thesis, University of Bern (2013).
- [12] I. Molchanov, *Theory of random sets*, Springer, 2005.
- [13] V. Grunert da Fonseca, C. M. Fonseca, The attainment-function approach to stochastic multiobjective optimizer assessment and comparison, in: *Experimental methods for the analysis of optimization algorithms*, Springer, 2010, pp. 103–130.
- [14] E. Zitzler, J. Knowles, L. Thiele, Quality assessment of Pareto set approximations, *Multiobjective Optimization* (2008) 373–404.
- [15] G. Matheron, Principles of geostatistics, *Economic geology* 58 (8) (1963) 1246–1266.
- [16] C. E. Rasmussen, C. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.
- [17] J. P. Kleijnen, E. Mehdad, Multivariate versus univariate kriging metamodels for multi-response simulation models, *European Journal of Operational Research* 236 (2) (2014) 573 – 582.
- [18] J. D. Svenson, T. J. Santner, Multiobjective optimization of expensive black-box functions via expected maximin improvement, *submitted* (2013).
- [19] M. L. Stein, *Interpolation of spatial data: some theory for kriging*, Springer, 1999.
- [20] J. Sacks, W. J. Welch, T. J. Mitchell, H. P. Wynn, Design and analysis of computer experiments, *Statistical science* 4 (4) (1989) 409–423.
- [21] M. S. Handcock, M. L. Stein, A bayesian analysis of kriging, *Technometrics* 35 (4) (1993) 403–410.
- [22] D. Ginsbourger, O. Roustant, N. Durrande, Invariances of random fields paths, with applications in Gaussian process regression, *arXiv preprint arXiv:1308.1359*.
- [23] F. Bachoc, Cross validation and maximum likelihood estimations of hyper-parameters of gaussian processes with model misspecification, *Computational Statistics & Data Analysis* 66 (2013) 55–69.
- [24] A. J. Keane, Statistical improvement criteria for use in multiobjective design optimization, *AIAA journal* 44 (4) (2006) 879–891.
- [25] M. Emmerich, K. Giannakoglou, B. Naujoks, Single-and multiobjective evolutionary optimization assisted by Gaussian random field metamodels, *Evolutionary Computation, IEEE Transactions on* 10 (4) (2006) 421–439.
- [26] T. Wagner, M. Emmerich, A. Deutz, W. Ponweiser, On expected-improvement criteria for model-based multi-objective optimization, *Parallel Problem Solving from Nature—PPSN XI* (2010) 718–727.
- [27] D. C. Bautista, A sequential design for approximating the pareto front using the expected pareto improvement function, Ph.D. thesis, The Ohio State University (2009).
- [28] A. G. Journel, Geostatistics for conditional simulation of ore bodies, *Economic Geology* 69 (5) (1974) 673–687.
- [29] M. Hoshiya, Kriging and conditional simulation of Gaussian field, *Journal of engineering mechanics* 121 (2) (1995) 181–186.
- [30] P. J. Diggle, P. J. Ribeiro, *Model-based geostatistics*, Springer, 2007.
- [31] O. Roustant, D. Ginsbourger, Y. Deville, DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization, *Journal of Statistical Software* 51 (1) (2012) 1–55.
- [32] J. Villemonteix, E. Vazquez, E. Walter, An informational approach to the global optimization of expensive-to-evaluate functions, *Journal of Global Optimization* 44 (4) (2009) 509–534.

- [33] C. M. Fonseca, V. Grunert da Fonseca, L. Paquete, Exploring the performance of stochastic multiobjective optimisers with the second-order attainment function, in: *Evolutionary Multi-Criterion Optimization*, Springer, 2005, pp. 250–264.
- [34] A. Auger, J. Bader, D. Brockhoff, E. Zitzler, Hypervolume-based multiobjective optimization: Theoretical foundations and practical implications, *Theoretical Computer Science* 425 (2012) 75–103.
- [35] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, V. Grunert da Fonseca, Performance assessment of multiobjective optimizers: An analysis and review, *Evolutionary Computation*, IEEE Transactions on 7 (2) (2003) 117–132.
- [36] J. M. Parr, Improvement criteria for constraint handling and multiobjective optimization, Ph.D. thesis, University of Southampton (2012).
- [37] E. Zitzler, K. Deb, L. Thiele, Comparison of multiobjective evolutionary algorithms: Empirical results, *Evolutionary computation* 8 (2) (2000) 173–195.
- [38] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *Evolutionary Computation*, IEEE Transactions on 6 (2) (2002) 182–197.
- [39] J. D. Svenson, Computer experiments: Multiobjective optimization and sensitivity analysis, Ph.D. thesis, The Ohio State University (2011).
- [40] C. Chevalier, X. Emery, D. Ginsbourger, Fast update of conditional simulation ensembles, Tech. rep., ReDICE Consortium (2014).
- [41] J. Oakley, Bayesian uncertainty analysis for complex computer codes, Ph.D. thesis, University of Sheffield (1999).
- [42] C. Chevalier, J. Bect, D. Ginsbourger, E. Vazquez, V. Picheny, Y. Richet, Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set, *Technometrics* doi:10.1080/00401706.2013.860918.